

Helix Systems -- Center for Information Technology
<http://helix.nih.gov>
 National Institutes of Health, Department of Health and Human Services
 DNABworks Web Site: <http://helixweb.nih.gov/dnaworks>

Send all correspondence to webtools@helix.nih.gov

Job started on 08/30/2013 at 16:00:10

Job name: JN083013HelicobacterpyloriACD

Output will be sent to jessonuno@utexas.edu

SEQUENCE 1: PROTEIN LENGTH = 416

```

1 MKKTFLIALVLATSLIGAENAKWDYKNKENGPHRWDKLHKDFEVCKSGKSQSPINIEHYY
61 HTQDKADLQFKYAASKPKAVFFTHHTLKASFEPTNHINRGHDYVLDNVHFHAPMEFLIN
121 NKTRPLSAHFVHKDAKGRLLVLAIGFEEGKENPNLDPILGIIQKKQNFKEVALDAFLPKS
181 INYYHFNGSLTAPPCTEGVAWFVIEEPLEVSQKLAIEIKRMKNSPNQRVPQPDYNTVII
241 KSSAKTRLGDSPGYVMSNIEFRQLTRGHSPSDEREARRVEEAGGQLFVIGGELRVNGVLN
301 LTRALGDVPGRPMISNEPETCQVPIESSDYLLACDGISDVFNERDLYQLVEAFANDYP
361 VEDYAELSRFICTKAIEAGSADNVSVVIGFLRPPQDVWKLKMHESDDESDVTDEE
  
```

CODON FREQUENCY TABLE: E. coli Class II

TTT F 0.291	TCT S 0.324	TAT Y 0.352	TGT C 0.389
TTC F 0.709	TCC S 0.266	TAC Y 0.648	TGC C 0.612
TTA L 0.034	TCA S 0.048	TAA X 0.630	TGA X 0.352
TTG L 0.055	TCG S 0.074	TAG X 0.076	TGG W 1.000
CTT L 0.056	CCT P 0.112	CAT H 0.298	CGT R 0.643
CTC L 0.080	CCC P 0.016	CAC H 0.702	CGC R 0.330
CTA L 0.008	CCA P 0.153	CAA Q 0.187	CGA R 0.011
CTG L 0.767	CCG P 0.719	CAG Q 0.814	CGG R 0.008
ATT I 0.335	ACT T 0.291	AAT N 0.173	AGT S 0.045
ATC I 0.659	ACC T 0.536	AAC N 0.828	AGC S 0.243
ATA I 0.006	ACA T 0.047	AAA K 0.786	AGA R 0.006
ATG M 1.000	ACG T 0.127	AAG K 0.215	AGG R 0.003
GTT V 0.398	GCT A 0.275	GAT D 0.461	GGT G 0.508
GTC V 0.135	GCC A 0.161	GAC D 0.540	GGC G 0.428
GTA V 0.200	GCA A 0.240	GAA E 0.754	GGA G 0.020
GTG V 0.268	GCG A 0.323	GAG E 0.247	GGG G 0.044

ACTIVE CODONS FOR SEQUENCE GENERATION

Residue	Codons	Active Codons	Min. Freq.	Max. Freq.
A Ala	4	4	0.161	0.323
C Cys	2	2	0.389	0.612
D Asp	2	2	0.461	0.540
E Glu	2	2	0.247	0.754
F Phe	2	2	0.291	0.709
G Gly	4	2	0.428	0.508
H His	2	2	0.298	0.702
I Ile	3	2	0.335	0.659
K Lys	2	2	0.215	0.786

L	Leu	6	2	0.080	0.767
M	Met	1	1	1.000	1.000
N	Asn	2	2	0.173	0.828
P	Pro	4	3	0.112	0.719
Q	Gln	2	2	0.187	0.814
R	Arg	6	2	0.330	0.643
S	Ser	6	3	0.243	0.324
T	Thr	4	3	0.127	0.536
V	Val	4	4	0.135	0.398
W	Trp	1	1	1.000	1.000
X	End	3	2	0.352	0.630
Y	Tyr	2	2	0.352	0.648

SEQUENCE PATTERNS TO BE SCREENED

PARAMETERS FOR TRIAL 1

Total Size Of Gene 1248 nt
 Protein Residues 416
 Mutatable Residues 406
 Fixed Nucleotides 30 nt
 Degenerate Nucleotides 0 nt
 Oligo Size 60 nt
 Annealing Temp 62 +/- 1°C
 Oligo Concentration 1.00E-7 M
 Sodium Concentration 5.00E-2 M
 Mg2+ Concentration 2.00E-3 M
 Codon Frequency Threshold .. 10%
 Repeat Threshold 8 nt
 Mispriming Threshold 8/18 (6 exact) nt

Too many sets of even overlaps -- increasing oligo length to 61

The DNA sequence # 1 is:

1 ATGAAAAAACCTTTCTGATCGCGCTCGTACTCGCGACTTCTCTCATCGGTGCGGAAAAAC
 61 GCGAAGTGGGATTACAAAAACAAGAGAACGGTCCGCATCGTTGGGATAAGCTGCACAAA
 121 GACTTCGAGGTATGCAAACTGGTAAGTCTCAGTCTCCGATCAACATCGAGCACTATTAC
 181 CATACCCAGGACAAAGCGGATCTGCAATTTAAGTACGCGGCGTCTAAACCTAAAGCAGTT
 241 TTCTTACCCACCACACCTCAAAGCATCTTTCGAACCGACCAACCACATCAATTACCGT
 301 GGTACGACTACGTTCTGGACAACGTTCACTTCCACGCGCCGATGGAGTTTCTCATCAAC
 361 AACAAGACCCGTCCTCTCAGCGCGCACTTCGTTCAAGGACGCGAAAGGTCGTCTGCTG
 421 GTAATCTTCTGCGAAAACTCGTCTCGGCGACTCCCGAGGTTACGTTATGTCTAATATCGAA
 481 GGTATCCAGAAAAACAGAATTTCAAAGAAGTTGCGCTCGACGCGTTCTGCCGAAATCC
 541 ATCAACTACTACCACTTCAACGGTTCTCTGACTGCCCTCCGTGCACCGAAGGTGTTGCG
 601 TGGTTTGTTATCGAGGAACCGTGGAAAGTTTCCGCCAAGCAGCTGGCGGAGATCAAAAAA
 661 CGTATGAAGAACTCTCAAACAGCGTCCGGTTCAGCCGCACTACAACACTGTTATCATC
 721 AAATCTTCTGCGAAAACTCGTCTCGGCGACTCCCGAGGTTACGTTATGTCTAATATCGAA
 781 TTCCGTCAACTCACTCGCGGTCACTCTCCGTCTGACGAACGTGAAGCGCGTCGTGTAGAA
 841 GAGGCGGGTGGTCAGCTGTTTCTGTTATCGGTGGTGAATCCGCGTCAACGGCGTTCTGAAC
 901 CTGACCCGTGCTCTGGGCGACGTGCCGGGTCGTCCGATGATCTCTAATGAACCAGAAACC
 961 TGTACGGTGCCTATTGAATCTTCCGACTACCTGGTTCTGCTCGCTTGCAGCGGCATTCT
 1021 GACGTGTTCAACGAACGCGACCTGTACCAACTGGTTGAAGCCTTTGCGAATGACTACCG
 1081 GTGGAAGACTACGCGGAAGTGTCTCGTTTCATCTGCACGAAAGCGATTGAAGCGGGTAGC
 1141 GCGGATAACGTTTCTGTCTGTCATTGGTTTCTGCGTCCGCCGCAAGACGTTTGGAACTG
 1201 ATGAAGCACGAATCCGACGACGAAGATAGCGACGTACCGACGAGGAG

The oligonucleotide assembly is:

```

-----
1      10      20      30      40      50      60
|      |      |      |      |      |      |
1 ---->
1 ATGAAAAAACCTTTCTGATCGCGCTCGTACTCGCGACTTCTCTCATCGGTGCGGAA
                                AGAGAGTAGCCACGCCTTTTG

    M K K T F L I A L V L A T S L I G A E N
    |      |      |      |      |      |      |

3 ---->
61 aaagagaacggtccgcatcggtgggataagctgcacaaa
CGCTTCACCCTAATGTTTTGTTTCTTGCCAGGCGTAG
                                <--- 2
    A K W D Y K N K E N G P H R W D K L H K
    |      |      |      |      |      |      |

5 ---->
121 gacttcgaggtatgcaaatctg                                GATCAACATCGAGCACTATTAC
ctgaagctccatacgttagaccattcagagtcagaggctagtgtgtagctcgtgataatg
                                <---
    D F E V C K S G K S Q S P I N I E H Y Y
    |      |      |      |      |      |      |

7
181 CATACCCAGGACAAAGCGGATCTGCAATTTAAGTACGCG                                tt
g                                CCTAGACGTTAAATTCATGCGCCGAGATTGGATTTCGTCAA
4
    H T Q D K A D L Q F K Y A A S K P K A V
    |      |      |      |      |      |      |

---->
241 ttcttcaccacaccctcaaagcatctttcgaaccgaccaaccacatcaattaccg
AAGAAGTGGGTGGTGTGG                                ctggttggtgtagttaatggca
                                <--- 6
    F F T H H T L K A S F E P T N H I N Y R
    |      |      |      |      |      |      |

9 ---->
301 ACGTTCACTTCCACGCGCCGATGGAGTTCCTCATCAAC
ccagtgctgatgcaagacctgttgcaagtgaaggtgcgc
                                <--- 8
    G H D Y V L D N V H F H A P M E F L I N
    |      |      |      |      |      |      |

11 ---->
361 AACAAAGCCCGTCCTCTCAGCGC                                tcgtctgctg
GGGCAGGAGAGTCGCGCGTGAAGCAAGTGTTCTGCGCTTTCAGCAGACGAC

    N K T R P L S A H F V H K D A K G R L L
    |      |      |      |      |      |      |

421 gtactggcgatcggtttcgaagaaggtaaagaaaatccgaacctcgatccg
CATGACCG                                ttttaggcttgagctaggctaggacctt
<--- 10
    V L A I G F E E G K E N P N L D P I L E
    |      |      |      |      |      |      |

```

13 ---->
 481 AGAAAAACAGAACTTCAAAGAAGTTGCGCTCGACGCGTTCTGCCGAAATCC
 ccataggtctttttgtcttgaagtttcttca GGACGGCTTTAGG
 <---- 12
 G I Q K K Q N F K E V A L D A F L P K S
 | | | | | | |
 15 ---->
 541 ATCAACTA gcccctccgtgcaccgaaggtgttgcg
 TAGTTGATGATGGTGAAGTTGCCAAGAGACTGACGGGAGGCACGTGG
 <---- 14
 I N Y Y H F N G S L T A P P C T E G V A
 | | | | | | |
 17 ---->
 601 tggtttgttatcgaggaaccgctggaagtttccg TCAAAAAA
 ttggcgaccttcaaaggcggttcgtcgaccgcctctagttttt
 W F V I E E P L E V S A K Q L A E I K K
 | | | | | | |
 16 ---->
 661 CGTATGAAGAACTCTCCAAACCAGCGTCCGGTTCAGCCGGACTACAACACTGT
 gcatacttcttgagagg TCGGCCTGATGTTGTGACAATAGTAG
 <---- 16
 R M K N S P N Q R P V Q P D Y N T V I I
 | | | | | | |
 19 ---->
 721 gtctcggcgactccccaggttacgttatgtctaatatcgaa
 TTTAGAAGACGCTTTTGAGCAGAGCCGCTGAGGGG
 <---- 18
 K S S A K T R L G D S P G Y V M S N I E
 | | | | | | |
 21 ---->
 781 ttccgtcaactcactcgcg CGCGTCGTGTAGAA
 gcagttgagtgagcgccagtgagaggcagactgcttgaccttgcgcagcacatctt
 <--
 F R Q L T R G H S P S D E R E A R R V E
 | | | | | | |
 20 ---->
 841 GAGGCGGGTGGTCAGCTGTTTCGTTATCGGTGGTGAACCTCCGCGTCAA
 ctcc ACCACTTGAGGCGCAGTTGCCGCAAGACTTG
 - 20
 E A G G Q L F V I G G E L R V N G V L N
 | | | | | | |
 23 ---->
 901 gcgacgtgccgggtcgtccgatgatctctaataaccagaaacc
 GACTGGGCACGAGACCCGCTGCACGGCCCA tgg
 <---- 22
 L T R A L G D V P G R P M I S N E P E T
 | | | | | | |
 25 ---->
 961 tgtcaggtgcctattga CGCTTGCGACGGCATTCT
 acagtcacggataacttagaaggctgatggaccaagacgagcgaacgctgccgtaaa
 <---- 24

C Q V P I E S S D Y L V L L A C D G I S
| | | | | | |

1021 GACGTGTTCAACGAACGCGACCTGTACCAACTGGTTGAAGCC
ACATGGTTGACCAACTTCGGAACGCTTACTGATGGGC

D V F N E R D L Y Q L V E A F A N D Y P
| | | | | | |

27 --->
1081 gaagactacgcggaactgtctcgtttcacatctgcacgaaagcgattgaagcgggtagc
CACCTTCTGATGCGCCTTGACAG ttcgcccacgcg

<--- 26
V E D Y A E L S R F I C T K A I E A G S
| | | | | | |

29 --->
1141 gcgg GTCCGCCGCAAGACGTTTGAAACTG
cgcctattgcaaagacagcagtaaccaaaggcgcggcggttctgca

<--- 28
A D N V S V V I G F L R P P Q D V W K L
| | | | | | |

1201 ATGAAGCACGAATCCGACGACGAAGATAGCGACGT
TGCTGCTTCTATCGCTGCAGTGGCTGCTCCTC

<--- 30
M K H E S D D E D S D V T D E E

The total codon usage score 0.000
The total length score 0.000
The total melting temperature score ... 0.000
The total repeat score 0.000
The total pattern score 0.000
The total mispriming score 0.000
The total AT content score 0.000
The total GC content score 0.000
The total fixed gap score N/A
The OVERALL score 0.000

DETAILED CODON FREQUENCY REPORT
[Codon, AA, Frequency, # of times used in coding sequence]

TTT F 0.29 4 TCT S 0.32 18 TAT Y 0.35 1 TGT C 0.39 1
TTC F 0.71 16 TCC S 0.27 5 TAC Y 0.65 13 TGC C 0.61 4
TTA L 0.03 0 TCA S 0.05 0 TAA X 0.63 0 TGA X 0.35 0
TTG L 0.06 0 TCG S 0.07 0 TAG X 0.08 0 TGG W 1.00 4

CTT L 0.06 0 CCT P 0.11 4 CAT H 0.30 2 CGT R 0.64 15
CTC L 0.08 12 CCC P 0.02 0 CAC H 0.70 13 CGC R 0.33 3
CTA L 0.01 0 CCA P 0.15 3 CAA Q 0.19 4 CGA R 0.01 0
CTG L 0.77 23 CCG P 0.72 17 CAG Q 0.81 9 CGG R 0.01 0

ATT I 0.34 4 ACT T 0.29 5 AAT N 0.17 5 AGT S 0.05 0
ATC I 0.66 18 ACC T 0.54 10 AAC N 0.83 19 AGC S 0.24 3
ATA I 0.01 0 ACA T 0.05 0 AAA K 0.79 22 AGA R 0.01 0
ATG M 1.00 6 ACG T 0.13 1 AAG K 0.22 9 AGG R 0.00 0

GTT V 0.40 17 GCT A 0.28 2 GAT D 0.46 6 GGT G 0.51 19
GTC V 0.14 4 GCC A 0.16 3 GAC D 0.54 21 GGC G 0.43 4

GTA V 0.20	4	GCA A 0.24	2	GAA E 0.75	25	GGA G 0.02	0
GTG V 0.27	4	GCG A 0.32	23	GAG E 0.25	9	GGG G 0.04	0

Frequency Range	Number of Codons
0% - 4%	0
5% - 9%	12
10% - 14%	9
15% - 19%	15
20% - 24%	27
25% - 29%	22
30% - 34%	48
35% - 39%	19
40% - 44%	4
45% - 49%	6
>= 50%	254

Total Codons Used = 416

Tm Range	# of Overlaps
<60	0
60	0
61	12
62	17
63	0
64	0
65	0
66	0
67	0
>=68	0

Tm Range = 1.9

Overlap Len Range	# of Oligos
<15	1
15	2
16	3
17	3
18	5
19	3
20	5
21	2
22	2
23	1
24	0
>=25	2

Lowest Overlap = 14

Length Range	# of Oligos
< 51	1
51- 52	0
53- 54	0
55- 56	0
57- 58	1
59- 60	0
61- 62	28
63- 64	0
65- 66	0
67- 68	0
69- 70	0
>= 71	0

Longest = 61

Sequence Patterns Screened (As Supplied By User)

None found

30 oligonucleotides need to be synthesized

```

1 ATGAAAAAACCTTTCTGATCGCGCTCGTACTCGCGACTTCTCTCATCGGTGCGGAA 57
2 GATGCGGACCGTTCTCTTTGTTTTGTAATCCCACTTCGCGTTTTCCGCACCGATGAGAGA 61
3 AAAGAGAACGGTCCGCATCGTTGGGATAAGCTGCACAAAGACTTCGAGGTATGCAAATCTG 61
4 GGTAATAGTGCTCGATGTTGATCGGAGACTGAGACTTACCAGATTTGCATACCTCGAAGTC 61
5 GATCAACATCGAGCACTATTACCATACCCAGGACAAAGCGGATCTGCAATTTAAGTACGCG 61
6 GGTGTGGTGGGTGAAGAAAACTGCTTTAGGTTTAGACGCCGCTACTTAAATTGCAGATCC 61
7 TTTTCTTCACCCACCACACCTCAAAGCATCTTTCGAACCGACCAACCACATCAATTACCG 61
8 CGCGTGGAAGTGAACGTTGTCCAGAACGTAGTCGTGACCACGGTAATTGATGTGGTTGGTC 61
9 ACGTTCACTTCCACGCGCCGATGGAGTTCTCTCATCAACAACAAGACCCGTCCTCTCAGCGC 61
10 GCCAGTACCAGCAGACGACCTTTTCGCGTCCTTGTGAACGAAGTGCAGCGCTGAGAGGACGGG 61
11 TCGTCTGCTGGTACTGCGCATCGGTTTCGAAGAAGGTAAGAAAAATCCGAACCTCGATCCG 61
12 ACTTCTTTGAAGTTCTGTTTTTCTGGATACCTTCCAGGATCGGATCGAGGTTTCGGATTTT 61
13 AGAAAAACAGAACTTCAAAGAAGTTGCGCTCGACGCGTTCTGCCGAAATCCATCAACTA 61
14 GGTGCACGGAGGGCAGTCAGAGAACCCTTGAAGTGGTAGTAGTTGATGGATTTTCGGCAGG 61
15 GCCCCCTCGTGACCGCAAGGTGTTGCGTGGTTTGTATCGAGGAACCGCTGGAAGTTTCCG 61
16 GGAGAGTTCTTCATACGTTTTTGTATCTCCGCCAGCTGCTTGGCGGAACTTCCAGCGGTT 61
17 TCAAAAAACGTATGAAGAACTCTCAAACCAGCGTCCGGTTCAGCCGGACTACAACACTGT 61
18 GGGGAGTCGCCGAGACGAGTTTTTCGAGAAGATTTGATGATAACAGTGTGTAGTCCGGCT 61
19 GTCTCGGCGACTCCCCAGGTTACGTTATGTCTAATATCGAATTCGGTCAACTCACTCGCGG 61
20 CCTCTTCTACACGACGCGCTTCACGTTCTGTCAGACGGAGAGTGACCGCGAGTGAGTTGACG 61
21 CGCGTCGTGTAGAAGAGCGGGTGGTCACTGTTTCGTTATCGGTGGTGAACCTCCGCGTCAA 61
22 ACCCGGCACGTCGCCAGAGCAGGGTCAGGTTTCAAGACGCCGTTGACGCGGAGTTTACCA 61
23 GCGACGTGCGGGTCTGTCGATGATCTTAATGAACCAGAACTGTCAAGTGCCTATTGA 61
24 AAATGCCGTGCGAAGCGAGCAGAAACCAAGTAGTCGGAAGATTCAATAGGCACCTGACAGGT 61
25 CGCTTGCGACGGCATTCTGACGTGTTCAACGAACGCGACCTGTACCAACTGGTTGAAGCC 61
26 GACAGTTCCGCGTAGTCTTCCACCGGGTAGTCATTGCGAAAGGCTTCAACCAAGTTGGTACA 61
27 GAAGACTACGCGGAAGTGTCTCGTTTTCATCTGCACGAAAGCGATTGAAGCGGGTAGCGCGG 61
28 ACGTCTTTCGCGCGGACGACGAGAAACCAATGACGACAGAAACGTTATCCGCGCTACCCGCTT 61
29 GTCCGCCGCAAGACGTTTGGAACTGATGAAGCACGAATCCGACGACGAAGATAGCGACGT 61
30 CTCCTCGTCGGTGACGTCGCTATCTTCGTCGT 32

```

FINAL SUMMARY FOR 1 SOLUTION

#	Tm	Len	Score	TmRange	Short	Long	#Olig	#Repeat	#Misprime
1	62	61	0.000	1.9	14	61	30	0	0

Helix Systems -- Center for Information Technology
<http://helix.nih.gov>
 National Institutes of Health, Department of Health and Human Services
 DNABWorks Web Site: <http://helixweb.nih.gov/dnaworks>